DOCUMENT RESUME

ED 428 085 TM 029 485

AUTHOR	Mertler, Craig A.
TITLE	Classroom Assessment Practices of Ohio Teachers.
SPONS AGENCY	Martha Holden Jennings Foundation, Cleveland, OH.
PUB DATE	1998-10-00
NOTE	21p.; Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Chicago, IL, October 14-17, 1998).
CONTRACT	A-107-97
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Classroom Techniques; *Educational Assessment; Educational Practices; Elementary Secondary Education; Performance Based Assessment; *Student Evaluation; *Teachers; Test Construction; *Test Use
IDENTIFIERS	*Alternative Assessment; *Ohio

ABSTRACT

A descriptive study was conducted to examine the current assessment practices of teachers in Ohio. The specific aim of the study was to gain an understanding of the extent to which teachers use traditional versus alternative forms of assessment techniques in their classrooms. Participants were 625 teachers from kindergarten through grade 12. The study resulted in a moderately thorough description of these teachers' assessment practices. It builds on previous research by incorporating descriptions of alternative assessment practices. As previous research found, teachers did not spend much time conducting statistical analyses of their assessment data. In addition, there are significant differences among teachers at different school levels and at differing levels of teaching experience with respect to their assessment practices, specifically in the use of traditional and alternative strategies and the frequency of use of specific item types. Recommendations include tailoring measurement courses to fit the needs of future classroom teachers. (Contains 5 tables and 10 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 428 085

CLASSROOM ASSESSMENT PRACTICES OF OHIO TEACHERS

Craig Mertler

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Craig A. Mertler, Ph.D.
Department of Educational Foundations & Inquiry
College of Education & Human Development
Bowling Green State University
Bowling Green, OH 43403

A Paper Presented at the 1998 Annual Meeting of the
Mid-Western Educational Research Association

Chicago, Illinois
October 14-17, 1998

BEST COPY AVAILABLE

TM029485

ABSTRACT

The general purpose of this descriptive study was to examine the current assessment practices of teachers in the state of Ohio. Specifically, the aim of this study was to gain an understanding of the extent to which teachers use traditional versus alternative forms of assessment techniques in their classrooms. The study resulted in a moderately thorough description of these teachers' assessment practices. It builds on previous research by incorporating descriptions of alternative assessment practices. Similar to previous research, it was determined that teachers do not spend much time conducting statistical analyses of their assessment data. Additionally, there are some significant differences between teachers at different school levels and of differing levels of teaching experience with respect to their assessment practices, specifically in their use of traditional and alternative strategies, and frequency of use of specific item types. Recommendations included tailoring measurement courses to fit the needs of future classroom teachers.

Note: This research was made possible by a grant from the Martha Holden Jennings Foundation, Cleveland, OH (Grant No. A-107-97). The author wishes to express his appreciation to the Foundation for its support.

INTRODUCTION

Assessment of student learning is a regular part of the school routine. A sizable amount of classroom time is devoted to the assessment of student learning. Since teachers must give even more time to the preparation and scoring of tests and other assessments, a substantial proportion of a teacher's day is devoted to issues surrounding student assessment. One could argue, then, that careful consideration of testing within formal teacher preparation programs is certainly warranted. If educators, particularly those in teacher preparation programs, are to help teachers use their student testing time efficiently and to be effective at it, more must be learned about how teachers perceive and use classroom tests and other forms of assessment (Gullickson, 1984).

For some time, there has been a perceived misalignment between what is taught to preservice teachers, in terms of assessment skills and techniques, and what inservice teachers actually practice in the schools (Farr & Griffin, 1973; Gullickson, 1986). Some have argued that measurement courses tend to overemphasize large-scale, standardized testing (Farr & Griffin, 1973; Stiggins & Bridgeford, 1985), as well as statistical analyses of classroom test data (Gullickson, 1986), neither of which serve teachers' primary measurement needs. It has been noted that teachers place much emphasis on non-test assessment and evaluation strategies (Gullickson, 1985). In his study, Gullickson (1984) reported that the average teacher did not perceive college measurement courses to be pertinent to his/her classroom testing needs and that most teachers learned how to test their students through their on-the-job experiences. We in higher education seem to have a limited understanding of the nature of assessment practices in K-12 classrooms (Stiggins & Bridgeford, 1985). From the perspective of the classroom teacher, this seems to imply a need for the reorientation of college instruction, with respect to measurement issues and concepts.

Several researchers have examined the traditional assessment practices of teachers and have arrived at somewhat similar conclusions. In their study, Stiggins and Bridgford (1985) discovered that about half of the teachers studied reported comfortable use of teacher-made objective tests. This finding held true across grade levels and subject areas. Marso (1985;

1987) arrived at the same conclusions for teachers in general, but did find several differences between elementary and secondary teachers. Secondary teachers tended to use more self-constructed tests rather than published tests; whereas, the opposite was true for elementary teachers, especially those in grades K-4. Similarly, others have found that the higher the grade level, the greater the tendency for teachers to use their own assessments (Stiggins & Bridgeford, 1985). Secondary teachers reported relatively more use of essay and problem-type items and less frequent use of completion and multiple-choice items than did elementary teachers (Marso, 1987). Marso (1985) also found that teachers perceived matching, multiple-choice, and completion type items as being most useful.

Marso & Pigge (1987) found no significant differences with respect to assessment practices based on school setting (urban, suburban, or rural) or age of teacher. However, subject area differences did exist. Teachers of mathematics reported more use of problem-type test items as compared to other subject areas, namely business, English, science, and social studies. Social studies teachers reported less frequent use of statistical analyses of test data, but more frequent use of essay items than did the other areas. Science teachers reported more frequent use of problem-type items than did English and social studies teachers. Similarly, Stiggins and Bridgeford (1985) found that teachers of mathematics and science tended to use their own objective tests slightly more that teachers of writing and speaking courses.

A final overriding theme in studies of teachers' traditional assessment practices is the infrequent use of statistical analyses of test data (Gullickson, 1986; Marso & Pigge, 1987; Marso & Pigge, 1988). This may be due to the fact that teachers are not convinced of the value of using statistical procedures to improve the quality of their tests or that they simply do not have a good grasp of statistical concepts and this discomfort may lead to a devaluing of their use.

The amount of attention focused on alternative assessment in the past fifteen years would lead one to believe that it has been a recent development in educational assessment. However, alternative forms of assessment have been around, and in use, for years. The recent attention paid to these forms of assessment is due more to mandated statewide assessments,

which have emphasized higher-order thinking and "real-world" application of reasoning skills. These factors have largely been responsible for bringing alternative assessment to the forefront of classroom assessment (Airasian, 1997).

However, many teachers remain uncomfortable with, or simply choose not to use, alternative forms of assessment with their students. The application of alternative assessment is appropriate at all grade levels, but they experience more frequent implementation for assessing student learning in performance-oriented areas (e.g., communication skills, psychomotor skills, concept acquisition, and affective skills) (Airasian, 1997). Their use is often seen in subjects such as art, music, physical education, writing, and even mathematics and science.

There is less research concerning teachers' alternative assessment practices. Stiggins and Bridgeford (1985) found that about three-fourths of the teachers studied reported some use of performance assessments in their classrooms. However, only about half of these teachers reported being comfortable with the use of these assessments.

Finally, Stiggins and Bridgeford (1985) discuss several relatively stable patterns across grade levels and subject areas. As grade level increases, teacher-made objective tests and structured performance assessments gradually increase in importance and classroom use. Teachers of mathematics and science tend to use their own objective tests whereas teachers of writing and speaking use performance assessments with greater frequencies.

PURPOSE

The general purpose of this study was to examine the current assessment practices of teachers in the state of Ohio. The researcher sought to explore how practicing teachers assess student performance with their students in their own classroom settings. Specifically, the aim of this research study was to gain an understanding of the extent to which teachers use traditional versus alternative forms of assessment techniques in their classrooms. The goal of the study was to describe the overall assessment practices of teachers, as well as examine

differences in practices based on gender, school level (elementary, middle, and high schools), school setting (urban, suburban, and rural), and years of experience.

The specific objectives for this study were:

(1) to determine the extent to which there were significant group differences with respect to the frequency of use of traditional and alternative assessment techniques based on the various levels of gender, school level, school setting, and years of teaching experience;

(2) to determine the extent to which there were significant group differences with respect to the use of statistical analyses of classroom assessment data based on school level and years of teaching experience;

(3) to determine the extent to which there were significant group differences with respect to the teachers' perceived level of preparation to assess student learning resulting from their undergraduate programs, as well as their current level of preparation, based on school level and years of teaching experience; and

(4) to determine the extent to which teachers ensure that their classroom assessments are both valid and reliable.

METHOD

Participants

Participants for the study consisted of 625 K-12 teachers. The sample consisted of 53% females and 47% males. The majority (42%) of teachers were from suburban settings, followed closely by those in rural (32%) and urban (25%) settings. Nearly half (47%) were teaching at the senior high level; just over one-fourth (26%) were teaching at the elementary level, followed closely by those teaching at the junior high/middle school level (25%). Twenty percent of the teachers had 26-30 years of teaching experience, followed by 21-25 years (19%), 6-10 years (17%), 1-5 years (13%), 16-20 years (13%), 11-15 years (11%), and 31-35 years (6%). Two teachers in the sample had 36 years or more of teaching experience.

Procedures

The researcher made use of resources available through the Ohio Department of Education in order to obtain a stratified random sample of K-12 teachers throughout the state of Ohio. The sample was stratified so that various subgroups in the population of K-12 teachers in the state were represented in the sample in the same proportion that they exist in the population. These subgroups of teachers included the following four categories: (1) female elementary, (2) female secondary, (3) male elementary, and (4) male secondary. A random sample of 3,000 teachers was obtained.

Each teacher received a packet containing a full-page cover letter, copy of the survey, and a self-addressed, postage-paid return envelope. They were instructed to return the survey within four weeks from the date appearing on the cover letter. Four weeks later, a follow-up reminder postcard was sent to those teachers who had not yet returned completed surveys. The final response rate was 21%.

Instrumentation and Analyses

An original survey instrument, the *Ohio Teacher Assessment Practices Survey*, was developed by the researcher for purposes of collecting the data. The literature was relied upon heavily in order to guide the development of the specific items appearing in the survey instrument. The instrument consisted of 47 items and included both scaled (forced-choice) and open-ended items. Teachers were asked to respond to items that addressed their use of traditional assessment and alternative assessment techniques, focusing on their frequency of use of these techniques. Additional items asked them to describe their comfort level with respect to assigning grades based on traditional versus alternative assessments, to describe any training they have received on the topic of student assessment, and to describe measures they take to ensure the validity and reliability of their classroom assessments. Finally, teachers were asked to indicate their gender, school setting, school level, years of experience, and subject area. Analyses of the resulting survey data included frequencies and percentages of responses, descriptive statistics, independent-samples *t*-tests, and one-way ANOVA's.

RESULTS

The results of the study will be presented for each objective as identified in the purposes of the study.

Objective 1: To determine the extent to which there were significant group differences with respect to the frequency of use of traditional and alternative assessment techniques based on the various levels of gender, school level, school setting, and years of teaching experience.

Participants were asked to respond to several items on a five-point Likert scale representing their frequency of use (i.e., 1 = "never", 2 = "not very often", 3 = "about half of the time", 4 = "most of the time", and 5 = "always"). For a portion of this set of analyses, two composite scores were obtained for each participant. A composite score for use of traditional assessment techniques (TRADCOMP) was comprised of nine items (with possible scores ranging from 9 to 45, where scores ranging from 9-16 = "never", 17-23 = "not very often", 24-30 = "about half of the time", 31-37 = "most of the time", and 38-45 = "always"). Similarly, a composite score for use of alternative assessment techniques (ALTCOMP) was comprised of eight items (with possible scores ranging from 8 to 40, where scores ranging from 8-14 = "never", 15-20 = "not very often", 21-26 = "about half of the time", 27-32 = "most of the time", and 33-40 = "always"). The results of the analyses of these two composite scores are summarized in Tables 1 and 2.

With respect to the frequency of use of traditional and alternative assessment techniques, no statistically significant gender differences existed for the use of traditional techniques; both males and females tended to use these techniques about half of the time. However, females reported that they use alternative assessment techniques about half of the time, which was significantly more frequently than their male counterparts who reported that they didn't use these techniques very often. Although teachers at the three school levels reported using traditional techniques about half of the time, elementary teachers use these assessments significantly less frequently than do both middle and high school teachers. Elementary teachers, who reported using alternative assessment techniques about half of the time, use alternative assessments significantly more frequently than do high school teachers,

who reported not using them very often. No statistically significant differences were found between teachers in urban, suburban, or rural schools with respect to their use of traditional assessments, but suburban teachers who reported using alternative assessment techniques about half of the time, use these assessments significantly more frequently than do rural teachers, who reported not using them very often. Similarly, no significant differences were found between teachers based on their years of teaching experience with respect to their use of traditional assessments (again, teachers reported using them about half of the time), but significant differences did exist for their use of alternative assessments. Teachers with 1-5 years of experience reported using alternative assessments about half of the time, significantly more frequently than their counterparts with 31-35 years of experience, who reported not using alternative techniques very often.

Insert Table 1 here

Insert Table 2 here

Elementary teachers reported that they don't use true/false items very often, which was significanlty less frequently than both middle and high school teachers, both of whom reported using these item types about half of the time. Elementary and high school teachers reported using multiple choice items significantly more frequently than middle school teachers. High school teachers use multiple choice items more frequently than both groups. No statistically significant differences were found for frequency of use of completion, short answer, or essay items; all teachers reported using these types of items about half of the time. These results are summarized in Table 3.

Insert Table 3 here

Elementary teachers reported using informal observations and questions most of the time, significantly more often that both middle and high school teachers, who reported using these informal techniques about half of the time. Similar results were obtained for the use of portfolios in that elementary teachers reported using them significantly more often than middle and high school teachers. No statistically significant differences existed with respect to the frequency of use of performance assessments; all teachers reported using these assessments about half of the time. These results are summarized in Table 4.

Insert Table 4 here

Objective 2: To determine the extent to which there were significant group differences with respect to the use of statistical analyses of classroom assessment data based on school level and years of teaching experience.

Although senior high teachers reported that they did not calculate means and standard deviations very often, they did so significantly more frequently than elementary teachers. There were no statistically significant differences between school levels when it comes to estimating reliability for tests or conducting item analyses; most teachers responded that they did not do either very often. No significant differences were found between differing years of experience for any of these variables.

No statistically significant differences existed between school levels with respect to estimating reliability for alternative assessments; most teachers also responded that they did not do this very often. No significant differences were found between differing years of experience for any of these variables. These results are summarized in Table 5.

Insert Table 5 here

Objective 3: To determine the extent to which there were significant group differences with respect to the teachers' perceived level of preparation to assess student learning

resulting from their undergraduate programs, as well as their current level of preparation, based on school level and years of teaching experience.

Teachers were asked to indicate their level of preparation in terms of assessing student performance that resulted from their undergraduate teacher education program as well as their current level of preparation. The median response from teachers regarding preparation resulting from their undergraduate programs was "slightly prepared" and the median response for their current level moderately improved to "somewhat prepared." This may imply that teachers tend to develop some assessment skills "on-the-job." There were no statistically significant group differences (i.e., by gender, school level, or years of experience) for undergraduate level of preparation, $F(2, 603) = .28$, $p > .05$, or current level of preparation, $F(2, 606) = .89$, $p > .05$.

Objective 4: To determine the extent to which teachers insure that their classroom assessments are both valid and reliable.

Teachers were asked to indicate how often they followed specific steps to insure that their assessments were both valid and reliable. One-fourth (25%) of the teachers responded that they followed specific steps to insure validity about half of the time or less; the median response was "most of the time." The teachers were also asked to indicate how often they believed that teachers in general followed specific steps to determine the validity of their assessments. Two-thirds (66%) of the teachers believed that teachers followed those steps about half of the time or less; the median response was "about half the time."

Similarly, nearly one-third (30%) of the teachers responded that they followed specific steps to insure reliability about half of the time or less; the median response was "most of the time." Two-thirds (66%) of the all teachers believed that teachers followed those steps about half of the time or less; the median response was "about half the time." With respect to insuring the validity and reliability of classroom assessments, the results seem to imply that these teachers believe that individually they are doing a better job than the majority of other teachers with whom they are familiar.

DISCUSSION/CONCLUSIONS

The purpose of this study was to examine the current assessment practices of teachers in the state of Ohio. Specifically, the aim of this research study was to gain an understanding of the extent to which teachers use traditional and alternative forms of assessment techniques in their classrooms. The author would be remiss if a strong limitation of this research study was left unidentified. The low rate of return (21%) may have been a by-product of the time of school year in which the surveys were distributed. Although on the surface this appears to be somewhat problematic, it was determined that the resultant sample size (n = 625) was sufficiently large to draw, at a minimum, preliminary conclusions.

This study was successful in that it resulted in a moderately thorough description of these teachers' assessment practices. It builds on previous research by incorporating descriptions of alternative assessment practices. Similar to previous research, it was determined that teachers do not spend much time conducting statistical analyses of their assessment data. Additionally, there are several statistically significant differences between teachers at different school levels, with differing years of experience, and different school locations with respect to their assessment practices. With respect to traditional assessment techniques, middle and high school teachers use these techniques more frequently than do elementary teachers. Specifically, teachers at the elementary level use true/false items less frequently than their counterparts at the middle and high school levels. Middle school teachers use multiple choice items less often than both of the remaining groups.

In general, teachers at the elementary level use alternative assessment techniques more frequently than high school teachers. Additionally, teachers in suburban schools use alternative assessments more often than teachers in rural settings. Teachers with fewer years of experience tend to use alternative assessments more frequently than teachers with 30 or more years of experience in the classroom. Elementary school teachers use informal observations and questions, as well as portfolios, more frequently than do middle and high school teachers.

By indicating that their current level of preparation in terms of assessing student learning is better than the preparation they received from their undergraduate programs, teachers may be implying that some classroom assessment skills are acquired on-the-job. Finally, with respect to insuring that assessments are both valid and reliable, teachers believe that they are doing a better job than most other teachers.

The general finding that assessment practices differ by school level (i.e., elementary, middle, high schools) perhaps implies that some restructuring of undergraduate teacher preparation measurement courses is warranted. There should be less focus on statistical concepts and increased attention paid to techniques of alternative assessment (including informal assessment strategies), which in many measurement courses, tend to be given cursory coverage. Although they still tend to use traditional slightly more often than alternative assessments, the teachers involved in this study indicated considerable use of alternative assessment techniques at all levels of K-12 education. Furthermore, since several differences were found between teachers at different levels of education (i.e., elementary, middle, and high schools), wherever possible the content of the courses should be altered to fit the needs of the level at which the preservice teacher will be teaching in the future. Only when measurement courses appropriately address the actual needs of classroom teachers will we have adequately prepared our teachers to assess their students' performance.

REFERENCES

Airasian, P.W. (1997). Classroom assessment. (3rd ed.). New York: McGraw-Hill.

Farr, R. & Griffin, M. (1973). Measurement gaps in teacher education. Journal of Research and Development in Education, 7(1), 19-28.

Gullickson, A.R. (1984). Teacher perspectives of their instructional use of tests. Journal of Educational Research, 77(4), 244-248.

Gullickson, A.R. (1985). Student evaluation techniques and their relationship to grade and curriculum. Journal of Educational Research, 79(2), 96-100.

Gullickson, A.R. (1986). Teacher education and teacher-perceived needs in educational measurement and evaluation. Journal of Educational Measurement, 23(4), 347-354.

Marso, R.N. (1985, October). Testing practices and test item preferences of classroom teachers. Paper presented at the annual meeting of the Mid-Western Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 268 145)

Marso, R.N. & Pigge, F.L. (1987, October). Teacher-made tests: Testing practices, cognitive demands, and item construction. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA. (ERIC Document Reproduction Service No. ED 298 174)

Marso, R.N. & Pigge, F.L. (1988, October). An analysis of teacher-made tests and testing: Classroom resources, guidelines, and practices. Paper presented at the annual meeting of the Mid-Western Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 291 781)

Marso, R.N. & Pigge, F.L. (1993, October). A summary of published research: Classroom teachers' and educators' attitudes toward and support of teacher-made testing. Paper presented at the annual meeting of the Mid-Western Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED 365 692)

Stiggins, R.J. & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271-286.

Table 1

Summary of *t*-Test Results for TRADCOMP and ALTCOMP by Gender

Dependent Variable	Group	*n*	*M*	*SD*	*t*-Statistic	prob *t*
TRADCOMP	Females	305	27.53	4.39	-1.39	.164
	Males	275	28.02	4.16		
ALTCOMP	Females	286	24.99	4.70	6.23	.000***
	Males	260	22.36	5.16		

*** significant at $p < .001$

Table 2

Summary of One-Way ANOVA Results for TRADCOMP and ALTCOMP by School Setting, School Level, and Years of Experience

Dependent Variable	Independent Variable	Levels	*n*	*M*	*SD*	*F*-Ratio	prob *F*
TRADCOMP	School Setting	Urban	138	27.44	4.15	.60	.550
		Suburban	250	27.94	4.45		
		Rural	189	27.75	4.18		
TRADCOMP	School Level	Elementary	155	26.59	4.87	8.61	.000***
		Middle	145	28.42	4.02		
		High	276	28.08	3.89		
TRADCOMP	Experience	1-5 years	80	28.11	4.44	.36	.925
		6-10 years	99	27.67	4.01		
		11-15 years	59	27.58	4.27		
		16-20 years	76	27.54	4.20		
		21-25 years	112	27.94	4.61		
		26-30 years	118	27.53	4.42		
		31-35 years	35	28.40	3.53		
		36 + years	2	26.00	.00		
ALTCOMP	School Setting	Urban	126	23.41	5.21	4.02	.019*
		Suburban	238	24.40	4.70		
		Rural	178	23.02	5.46		
ALTCOMP	School Level	Elementary	144	24.84	4.44	5.06	.007**
		Middle	143	23.46	5.04		
		High	255	23.21	5.35		
ALTCOMP	Experience	1-5 years	72	25.40	4.45	3.14	.003**
		6-10 years	97	23.34	5.39		
		11-15 years	59	24.22	4.09		
		16-20 years	67	23.93	5.10		
		21-25 years	101	24.36	5.48		
		26-30 years	113	22.85	5.23		
		31-35 years	36	21.44	4.23		
		36 + years	2	22.00	4.24		

* significant at $p < .05$
** significant at $p < .01$
*** significant at $p < .001$

Table 3

Summary of One-Way ANOVA Results for Traditional Item Types by School Level

Dependent Variable	Independent Variable	Levels	*n*	*M*	*SD*	*F*-Ratio	prob *F*
True/False	School Level	Elementary	161	2.11	.97	5.64	.004**
		Middle	153	2.49	1.14		
		High	290	2.39	1.05		
Multiple Choice	School Level	Elementary	163	2.64	.98	14.23	.000***
		Middle	153	2.25	1.06		
		High	290	2.93	1.04		
Completion	School Level	Elementary	161	3.04	.99	.12	.884
		Middle	154	3.06	1.08		
		High	287	3.01	1.08		
Short Answer	School Level	Elementary	162	3.15	1.08	.97	.379
		Middle	154	3.23	1.01		
		High	289	3.30	1.05		
Essay	School Level	Elementary	161	2.50	1.25	2.74	.065
		Middle	151	2.76	1.23		
		High	290	2.78	1.28		

** significant at $p < .01$
*** significant at $p < .001$

Table 4

Summary of One-Way ANOVA Results for Alternative Assessments by School Level

Dependent Variable	Independent Variable	Levels	*n*	*M*	*SD*	*F*-Ratio	prob *F*
Informal Observations/ Questions	School Level	Elementary	163	3.61	.91	11.31	.000***
		Middle	154	3.16	1.06		
		High	291	3.13	1.14		
Portfolios	School Level	Elementary	160	2.71	1.24	13.18	.000***
		Middle	155	2.18	1.16		
		High	290	2.13	1.18		
Exhibitions/ Recitals	School Level	Elementary	162	2.81	1.08	.143	.867
		Middle	155	2.81	1.09		
		High	290	2.76	1.17		
Performance Assessments	School Level	Elementary	159	3.06	1.08	.189	.828
		Middle	152	3.13	1.08		
		High	286	3.13	1.13		

*** significant at $p < .001$

Table 5

Summary of One-Way ANOVA Results for Statistical Analyses of Classroom Assessments by School Level

Dependent Variable	Independent Variable	Levels	*n*	*M*	*SD*	*F*-Ratio	prob *F*
Traditional - Means & SD's	School Level	Elementary	163	1.77	1.06	4.14	.016*
		Middle	154	2.06	1.23		
		High	290	2.09	1.21		
Traditional - Estimate Reliability	School Level	Elementary	163	2.17	1.26	2.41	.091
		Middle	154	2.44	1.23		
		High	289	2.42	1.27		
Traditional - Item Analyses	School Level	Elementary	162	2.53	1.29	.571	.566
		Middle	155	2.55	1.26		
		High	287	2.44	1.22		
Alternative - Estimate Reliability	School Level	Elementary	153	2.35	1.19	.116	.891
		Middle	152	2.36	1.15		
		High	276	2.30	1.17		

* significant at $p < .05$